LIETUVOS BANKAS
EUROSISTEMA

Sectoral Production and Diffusion Index Forecasts for Output in Lithuania

Discussion Paper Series

No. 12 / 2019

ISSN 2345-0835 (online)

Discussion Paper Series

No. 12 / 2019

Soroosh Soofi-Siavash

(Bank of Lithuania, Kaunas University of Technology)[†]

Kristina Barauskaite

(Bank of Lithuania, ISM University of Management and Economics)[‡]

May 2019[§]

[†] Bank of Lithuania and Kaunas University of Technology, ssoofi@lb.lt.

[‡] Bank of Lithuania and ISM University of Management and Economics, kgriskeviciene@lb.lt.

[§] We thank seminar participants at the Bank of Lithuania and an anonymous referee for constructive comments.

Gedimino pr. 6, LT-01103 Vilnius

www. lb.lt

Discussion Papers describe research in progress by the author(s) and are published to stimulate discussion and critical comments.

The series is managed by Applied Macroeconomic Research Division of Economics Department and the Center for Excellence in Finance and Economic Research.

ABSTRACT

In this paper, we develop and describe quarterly data on disaggregated sectors in Lithuania which covers the period 1998-2018. The data is useful for empirical studies requiring panels with a large number of time observations and a large number of cross-sectional units. We follow the NACE2 level of disaggregation in developing our data, thus allowing us to combine the data with world input-output tables which we in turn use to identify the hubs and the main importing and exporting sectors within the economy. The data is then used for forecasting the growth rate of GDP. There is a substantial increase in the degree of covariation among sectoral production growth rates, which is observed using a split sample around 2008. When we apply factor methods, we find that this strong covariation can be explained by a few factors which are closely correlated to the growth of the retail and wholesale sectors. For GDP growth, the forecasts we consider are the diffusion index forecasts produced using a few indexes that summarize sectoral data, and the forecasts produced using the production growth of selected hubs and importing and exporting sectors. We find that the diffusion indexes and the production growth of sectors that make heavy use of imported inputs in their production have interesting forecasting power for the growth rate of GDP in the 2006-2011 and 2012-2018 periods.

Keywords: factor model, forecasting, input-output linkages, intersectoral networks.

JEL codes: E27, E37, C3, C67.

1 Introduction

In this paper, we develop a quarterly dataset on disaggregated sectors in Lithuania that is useful for empirical studies requiring panels with a large number of time observations and a large number of cross-sectional units. Then, we use the sectoral data for forecasting the growth rate of the GDP. Decisions taken by diverse entities, such as policymakers, consumers, and investors, are made on the basis of macroeconomic forecasts, the accuracy of which have important consequences. The forecasts are likely based on information sets which include a large number of disaggrageted variables. In the past decade, the number of forecasting studies using large disaggregated datasets has grown rapidly. This growth was made possible by the increasing availability of macroeconomic datasets together with the development of suitable tools for analysis of the data. One such tool is factor models, which relate a large portion of covariations in the data as if that stems from a few factors. Empirical evidence shows that factor models fit sectoral data well. See for example, Long and Plosser (1987) and Forni and Reichlin (1998) for the application of factor methods to sectoral employment and production data. In addition, an increasing number of studies use models that involve factor structures for macroeconomic forecasting. Given that there is a large degree of covariation in the disaggregated data on production, employment, price, etc., a large number of variables can be summarized by a few indexes, which are in turn used for forecasting. The indexes are factor-based, and following Stock and Watson (2002b), factor-based forecasts are referred to as *diffusion index forecasts*. Eickmeier and Ziegler (2008) provide a survey of the studies concerning diffusion index forecasting, and Boivin and Ng (2005) assess the econometric methods used in these studies.

Concerns about the sources of aggregate fluctuations from dissaggregated shocks have stimulated theoretical economic models which abstract from common shocks and rely exclusively on the role played by small idiosyncratic shocks, either arising for firms or detailed disaggregated sectors. Acemoglu et al. (2012) argue that microeconomic idiosyncratic shocks can lead to aggregate fluctuations due to the existing intersectoral input-output linkages and their unbalanced structure.[1] These results are supported by other authors, including Carvalho (2014), Atalay (2017), and Caliendo, Parro, Rossi-Hansberg, and Sarte (2018).

The goal of this paper is twofold. First, it seeks to describe a detailed quarterly dataset on sectors in Lithuania. The data includes 60 sectors (classified according to the NACE2 level) and covers the 1998-2018 period. One advantage of this level of sectoral disaggregation is the possibility of matching the sectoral time series variables with the available annual world input-output tables. This detailed data can serve as a starting point for empirical studies which use large panels and factor models, and it allows us to identify three types of sectors: (i) hub sectors, which play an important role in supplying products to other sectors, (ii) sectors which export abroad the most, and (iii) sectors that make heavy use of imported inputs. To the best of our knowledge, this paper is the first attempt to describe such a dataset for Lithuania and use it to carry out a systematic forecasting exercise.

Our second goal is to make twenty-plus forecasts for GDP growth and compare their performance with that of a standard benchmark model. A number of forecasts are factor-based, and others are guided by literature that tracks the effects of economic activity drivers of individual sectors on aggregate macroeconomic performance. In particular, we use diffusion indexes, the growth rate of production, sales, and turnovers of important individual sectors, as leading indicators for forecasting the growth rate of GDP. The ultimate test of forecasts is their out-of-sample performance, which shows how the forecasts perform in real time. Thus, we carry out pseudo out-of-sample forecasting exercises for forecast evaluation in different subsamples.

[1] Acemoglu et al. (2012) questions the famous diversification argument of Lucas (1977) that microeconomic shocks would average out and are less likely to have a significant impact on macroeconomic level. Gabaix (2011) also shows that if firms are of different sizes and contribute unequally to the final aggregate output the diversification argument breaks down.

Our finding is twofold. First, using factor methods, we find that there is a strong covariation among the sectors in the post-2008 period. This covariation can be explained by a few factors which are closely correlated with growth rates of the retail and wholesale sectors. Second, based on the pseudo out-of-sample forecasting results, we find that the diffusion indexes and production growth of the sectors that make heavy use of imported inputs have an interesting forecasting power for the growth rate of GDP in both the 2006-2011 and 2012-2018 periods.

The organization of the paper is as follows. Section 2 describes the data construction and the hub sectors selection procedures. Section 3 provides a factor analysis of sectoral data. In Section 4, we begin with a graphical analysis and move on to a quantitative analysis of the forecasts for GDP growth. Section 5 concludes.

2 Details of the Data

2.1 Production, Turnover and Sales

Data on production, turnover, and sales are obtained from Statistics Lithuania. We use quarterly indexes of manufacturing and construction production, turnover of retail and wholesale, and sales of services sectors. There are 60 sectors in total corresponding to the NACE2 disaggregation level, and the dataset covers the 1998Q1-2018Q1 period. All the series are seasonally adjusted and transformed to growth rates, at an annual rate. In particular, let X_{it} denote the value for sector i at date t. The growth rate is calculated as $x_{it} = 400 \times ln\left(X_{i,t}/X_{i,t-1}\right)$. The sectors are listed in Tables 7-9 in the appendix.

The lower panel of Figure 1 shows four-quarter averages of the annualized growth rates for all the sectors, defined as $\left(\sum_{l=0}^{3} x_{it-l}\right)/4 = 100 \times ln\left(X_{i,t}/X_{i,t-3}\right)$. The sectoral rates show large differences among themselves, with these differences evolving over time. To summarize the distributions of the sectoral rates in each period, the upper and middle

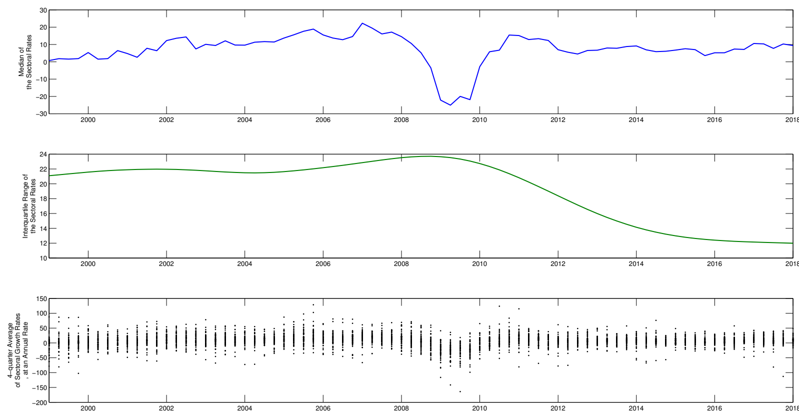

Figure 1: Four-quarter averages of the growth rates for all 60 sectors, at an annual rate, are shown in the lower panel. To summarize the marginal distributions (distribution in each period), the upper and middle panels show the median and the HP trend of the interquartile rage of sectoral rates, respectively.

panels of Figure 1 show the median of the growth rates and a smooth measure of sectoral dispersion, which is the Hodrick-Prescott smoothing of the quarterly interquartile range of sectoral rates obtained using the smoothing parameter 1600. We see that the median closely follows the economic cycles and falls to around -25 percent in 2009. Furthermore, the dispersion measure marks a steady decline in cross-sector differences in sectoral rates starting in 2009, which becomes flattened towards the end of the sample period. Over this period, we see that the HP trend of the intequartile rage decreases from a number above 20 percent to around 12 percent.

2.2 Intersectoral Data

We use the yearly Lithuanian intersectoral transactions tables from the world input-output (WIOD) database for the 2006-2014 period. This period is chosen because it reflects the out-of-sample period used below to investigate accuracy of the forecasts. As will be discussed below, some of the forecasts are made by using the growth rates of important sectors, detected from the input-output tables, as leading indicators.

The hubs	
1.	Electricity, gas, steam and air conditioning supply
2.	Wholesale trade. except of motor vehicles and motorcycles
3.	Land transport and transport via pipelines
4.	Warehousing and support activities for transportation
5.	Real estate activities
The sectors heavily using imported inputs	
1.	Coke and refined petroleum products
2.	Chemical products
3.	Rubber and plastic products
4.	Electrical equipment
The sectors exporting the most	
1.	Coke and refined petroleum products
2.	Wholesale trade, except of motor vehicles and motorcycles
3.	Land transport and transport via pipelines
4.	Manufacture of food products, beverages and tobacco products

Table 1: This tables presents five hubs (sectors which play an important role in supplying intermediate products to other sectors), four sectors that make heavy use of imported inputs, as well as four sectors which export the most among the Lithuanian sectors, at the NACE2 disaggregation level. The selection of these sectors is based on the 2006-2014 world input-output tables.

The world input-output tables provide intersectoral transactions at basic prices for the economy disaggregated to the 54 sectors. These tables are used to identify the sectors that are most important for supplying products to other sectors ("hubs").[2] The hub sectors in Lithuania are *Electricity, gas, steam and air conditioning supply*, *Wholesale trade. except of motor vehicles and motorcycles*, *Land transport and transport via pipelines*, *Warehousing and support activities for transportation*, and *Real estate activities*. Furthermore, since the WIOD database separates both national and international suppliers of intermediate

[2]To be more precise, from the input-output tables, we construct technical coefficients a_{ij}´s which show the flow of products from industrial sectors (i´s), to the same sector and all others (j´s). By adding up these weights, we get two main intersectoral network measures: total sectoral in-degrees and out-degrees. In-degrees capture the amount of intermediate goods a particular sector needs to purchase from all sectors in the economy while producing its output. At the same time, total out-degrees measure the amount of its final output sector sells as intermediates to all sectors in the economy. In turn, a hub sector is a sector whose out-degree is larger than the out-degrees of the others. For a more detailed explanation, refer to Acemoglu et al. (2012) and Constantinescu and Barauskaite (2018).

goods, it allows us to distinguish sectors which use the most imported intermediate inputs. In Lithuania, for the period 2006-2014, the highest importers of intermediate goods were *Coke and refined petroleum products*, *Chemical products*, *Rubber and plastic products*, and *Electrical equipment*. Additionally, we also select four main sectors which export the most, including the *Wholesale trade, except of motor vehicles and motorcycles* and *Land transport and transport via pipeline* sectors. Table 1 lists the selected sectors. The combination of these sectors in all three categories remains the same throughout the 2006-2014 period. This means, for example, that the *Wholesale trade, except of motor vehicles and motorcycles* sector remains among the hubs from one year to another.

3 Factor Analysis

3.1 Model

We assume that sectoral growth rates follow a factor model

$$x_{it} = \sum_{j=1}^{r} \lambda_{ij} f_{jt} + e_{it}, \tag{1}$$

where the growth rate of sector i, x_{it} ($i = 1,\ldots,60$), is related to r factors, f_{jt} ($j = 1,\ldots,r$), and an idiosyncratic error, e_{it}. The factor loadings are denoted by λ_{ij}s. The crucial assumption of the model is that a large portion of the covariation among sectoral rates stems from r factors. These factors basically capture the shocks which affect a large number of sectors.

There is a considerable empirical evidence that factor models fit macroeconomic data well (e.g. Forni and Reichlin 1998, Stock and Watson 2016). Furthermore, because a large number of predictors are summarized by a few factors, factor models have been developed as powerful tool for macroeconomic forecasting (e.g. Stock and Watson 2002a,b, and Boivin and Ng 2005).

Because the number of factors and the factors themselves are unobserved, we need to estimate them. Let T and N be the number of observations and the number of sectors, respectively (in our application, we have $N = 60$ and $T = 81$). For factor estimation, here we use the principal component method.[3] Furthermore, the number of factors, r, can be estimated using the Bai and Ng (2002)'s information criteria. In particular, consider

$$\mathrm{IC}(r) = \ln MSE + r\mathrm{g}(N,T), \tag{2}$$

where the first term is the mean squared error evaluated after including r principal components into the model, and $\mathrm{g}(N,T)$ is a penalty function proposed by Bai and Ng (2002). This follows the idea of AIC or BIC, and $\hat{r} = argmin_{0\leq r\leq r_{max}} \mathrm{IC}(r)$ estimates the number of factors. We here use the IC2 penalty function since it performs the best in simulation studies calibrated to macroeconomic data.

3.2 Statistics for Importance of the Factors

We begin with estimating the number of factors in the full sample, as well as that in two split subsamples around 2008. As is seen in the first part of Table 2, the Bai-Ng IC2 selects one factor for the full sample (1998-2007) and the second sample period (2008-2018), while selects zero factors in the first sample period (1998-2007). As is presented in the second part of Table 2, we compare the explanatory power of a single-factor model with the models including a higher number of factors. It suggests that one factor explains around 17 percent of the variation in the full sample, 13 percent of the variation in the first sample period, and 25 percent of those in the second sample period. At the same time,

[3] To be precise, let X be the $T \times N$ matrix of the growth rates. Furthermore, let F be the $T \times r$ matrix of the factors, and let Λ be the $N \times r$ matrix of factor loadings. Then, the principal component estimate of the factors, $\hat{F}$, is given by the first r eigenvectors of $XX'/(NT)$. In addition, under the normalization that $F'F/T = I_r$ where I_r is an identity matrix, the principal component estimate of the matrix of factor loadings is given by $\hat{\Lambda} = X'\hat{F}/T$.

adding the second factor increases the value of trace R^2 to 25 percent for the full sample, and to 24 percent and 34 percent for the first and the second periods accordingly.

Statistics for importance of factors						
	1998-2007		**2008-2018**		**Full sample**	
Bai-Ng IC2	0		1		1	
	1998-2007		**2008-2018**		**Full sample**	
Number of factors	**Trace R^2**	**Marginal trace R^2**	**Trace R^2**	**Marginal trace R^2**	**Trace R^2**	**Marginal trace R^2**
1	0.13	0.13	0.25	0.25	0.17	0.17
2	0.24	0.11	0.34	0.09	0.25	0.08
3	0.34	0.10	0.40	0.06	0.31	0.07
4	0.41	0.07	0.46	0.06	0.36	0.05
5	0.48	0.07	0.51	0.05	0.41	0.05

Table 2: Statistics for importance of the factors. The trace R^2 value shows the fraction of the variation in all the 60 sectoral growth rates explained by the number of factors in each row. The marginal R^2 values shows how much each factor adds to the explanatory power of the model.

Figures 2 and 3 show the marginal trace R^2 values for all the sectors explained by the first and second factors accordingly. We see that one factor explains most of the variation - up to 70 percent - of the growth rates of the construction and retail and wholesale sectors, though it has a smaller explanatory power for other sectors that varies between 0 and 45 percent. In addition, the second factor increases the R^2 values by 25 percent the most, particularly for some manufacturing and services sectors. In what follows, we use a two-factor model since, as compared to the single-factor model, it has additional explanatory power and captures the movements of a wider range of sectors. Moreover, the second factor has useful forecasting power for GDP growth in Lithuania, as discussed in Section 4.

Table 3 lists eight sectors with larger R^2 values which are explained by two factors. The first four sectors are retail and wholesale trade sectors, including *Retail trade, except for motor vehicles and motorcycles* ($R^2 = 0.7$) and *Wholesale and retail trade and repair of motor vehicles and motorcycles* ($R^2 = 0.63$). The remaining four sectors are manufacturing and services sectors including *Food and beverage service activities* ($R^2 = 0.61$), *Manufacture of fabricated metal products, except machinery and equipment* ($R^2 = 0.49$), and *Accomodation* ($R^2 = 0.49$).

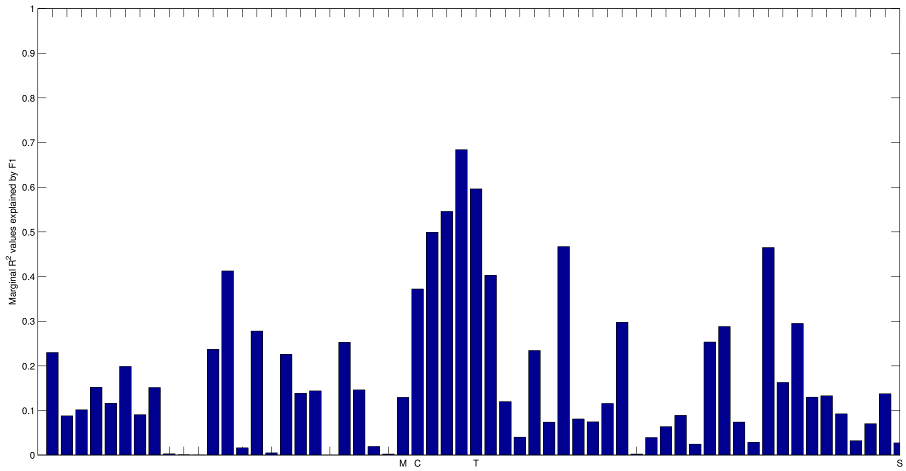

Figure 2: Marginal R^2 values explained by the first factor. On the x-axis from left to right, we have 26 manufacturing sectors (M), 1 construction sector (C), 4 retail and wholesale sectors (T) and 29 services sectors (S).

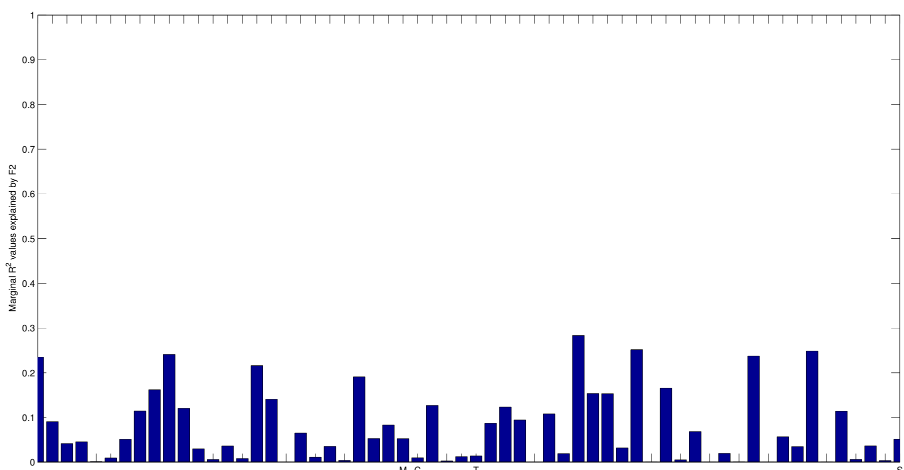

Figure 3: Marginal R^2 values explained by the second factor. On the x-axis from left to right, we have 26 manufacturing sectors (M), 1 construction sector (C), 4 retail and wholesale sectors (T) and 29 services sectors (S).

Sectors	R^2values explained by two factors
Retail trade, except for motor vehicles and motorcycles	0.70
Wholesale and retail trade and repair of motor vehicles and motorcycles	0.63
Food and beverage service activities	0.61
Wholesale trade, except for motor vehicles and motorcycles	0.55
Manufacture of fabricated metal products, except machinery and equipment	0.49
Land transport and transport via pipelines	0.49
Accommodation	0.49
Rental and leasing activities	0.47

Table 3: Eight sectors ordered by their R^2 values which are explained by two factors

What fraction of the variation of the growth in GDP, industrial production (IP), and sales of services can be explained by the sectoral growth rates most correlated to the factors, and what fraction of those can be explained by the growth rates of the hubs and importing and exporting sectors? Table 4 shows the R^2 values for four (eight) sectors (the growth rates of these sectors are most correlated to the factors and they are listed in Table 3), and the values for the hubs, importing and exporting sectors (these sectors are listed Table 1). We see that two factors explain 74 percent of the variation in GDP growth, 39 percent of the variation in IP growth, and 62 percent of the variation in the growth rate of sales of services, respectively. These numbers are 64 percent, 20 percent, and 52 percent for four sectors, respectively, and they increase to 74 percent, 31 percent, and 79 percent, if we include the growth rates of eights sectors. Furthermore, the growth rates of the hubs explain 84 percent of the growth variation in sales of services. The importing sectors growth rates explain around 81 percent of the growth variation in IP, and the rates of exporting sectors explain 56 percent of the variation in GDP growth, 78 percent of the variation in IP growth, and 73 percent of variation in the growth rates of sales of services, respectively. All the importing sectors are among the manufacturing sectors, which may explain why the importing sectors have high explanatory power for IP growth. In contrast, the exporting sectors belong to manufacturing and services sectors, which suggests that their growth rates have a high explanatory power for the growth in IP and sales of services, as was found according to the R^2 values.

R^2 values explained by individual sectors						
Variable	**Two factors**	**Four sec.**	**Eight sec.**	**Hubs**	**Importing sec.**	**Exporting sec.**
GDP	0.74	0.64	0.74	0.43	0.27	0.56
IP	0.39	0.20	0.31	0.22	0.81	0.78
Sales of services	0.62	0.52	0.79	0.84	0.15	0.73

Table 4: As a reference, this table shows the fraction of the variation in the growth in GDP, IP, and sales of services that is explained by two factors. In addition, the entries for the four (eight) sectors give the fraction of the variation explained by the rates of the first four (eight) sectors that are the most correlated with the factors (these sectors are presented in Table 3). The remaining columns show the R^2 values obtained by regressing on the rates of hubs, importing and exporting sectors (these sectors are presented in Table 1).

In sum, our results show that a large portion of the variation in GDP growth is explained by the contemporaneous variation in a few factors and the growth rates of a few sectors. An important question thus arises: Do the factors and production of individual sectors have the power to explain future movement in GDP growth? In the next section, we turn to forecasting GDP growth.

4 Forecasting GDP Growth

4.1 Constructing Forecasts

The GDP growth forecasts are produced by regressing the future values of GDP growth on the current and past values of growth, and the current and past values of a vector of other predictors, denoted by z_t. Let $y^h_{t+h} = (400/h)\ln(GDP_{t+h}/GDP_t)$ be the average GDP growth rate over next h quarters, at an annual rate, where h is a forecast horizon. Forecasts are produced using the model

$$y^h_{t+h} = c + \sum_{i=0}^{p-1} b_i y^h_{t-i} + \sum_{i=0}^{q-1} a_i' z_{t-i} + u^h_{t+h}, \tag{3}$$

where c, b_is and a_is are coefficients, and u^h_{t+h} is a forecast error.

Forecasts are produced using the data available prior to the forecasting dates. This means that, for example, for the two-quarter-ahead forecast for 2007Q2, we just use data available prior 2006Q4, or for the two-quarter-ahead forecast for 2007Q3, we just use data available prior 2007Q1. In this section, our goal is to test forecasting performance of our models in real time. True real time forecasts are produced without using the future values of the series. By using data available prior to each forecasting date and comparing the results with the true future value of GDP growth, we test how our forecasts performed in real time.

In the equation (3), we include dummy variables for the quarters 2009Q1-2009Q4 to remove the influence of the financial crisis from the coefficient estimates and the forecasts. The lag order is selected using BIC with $0 \le p,q \le 3$. When the lag orders p and q are selected to be zero, the regression contains only the intercept c.

We consider the 2006-2011 and 2012-2018 periods. In the recursive estimation of the forecasts, as discussed above, the first forecast is made using the observations available before 2005Q4, which means that the in-sample period includes thirty two observations. Considering the 2006-2011 and 2012-2018 periods splits the rest of the sample period into halves. We use the first subsample period to investigate the accuracy of the forecasts made in “bad times”, which include the crisis period, and use the second subsample period to investigate the accuracy of the forecasts made in “good times”, as compared to the first period.

We produce a number of different forecasts. We consider the diffusion index forecasts. Suppose that the sectoral growth rates follow the factor model (1). The factor estimates summarize the growth of sectors and are used as predictors for forecasting GDP growth. Let $\hat{f}_{1t}$ and $\hat{f}_{2t}$ be the first and the second principal components extracted from all the sectoral rates. In turn, we produce the diffusion index forecasts using $\hat{f}_{1t}$ and/or $\hat{f}_{2t}$ as predictors, z_ts, in (3). An important question now arises: do we need all the disaggregated series for forecasting GDP growth where the forecasts are produced using the factor-based indexes? As pointed by Boivin and Ng (2006), the answer to this question might be no. A possible reason is that the factors which are relevant for forecasting might underlie a subset of sectors, and in turn, the principal components extracted from all the growth rates might provide noisy estimates of the factors. To address this concern, we also consider the Bai and Ng (2008)'s modification of diffusion index forecasts. To do so, for each forecast, we first find the x_{it}s most correlated with the future GDP growth y^h_{t+h}. We next extract principal components from the preselected set. An advantage of this modification is that the forecast variable is considered when selecting the most relevant predictors, and the principal components are extracted from a selected set that is aimed for forecasting y^h_{t+h}. In our application, we find 20 most correlated sectoral rates in the first step. For selection of predictors, we use the Efron, Hastie, Johnstone, and Tibshirani (2004)'s least angle regression (LAR). LAR provides a computationally efficient approach to solve the LASSO.[4] See Bai and Ng (2008) for an application of LAR and other alternative methods for modification of diffusion index forecasts.

Furthermore, we consider the forecasts produced using the growth rate of individual sectors. The sectors considered are: the sectors whose growth rate is most correlated with diffusion indexes; the hub sectors, which are sectors that play an important role is supplying products to other sectors; the sectors that make heavy use of imported inputs; and the sectors that export the most. Among these rankings, there are some sectors which overlap. In total, thirteen sectors are considered. A complete list of these sectors is presented in Tables 1 and 3.

[4]The LASSO is a shrinkage method that solves a penalized least squares problem using an L_1 penalty. More particularly, consider the regression $y^h_{t+h} = \beta_0 + \sum_i \beta_i x_{it} + e^h_{t+h}$ in our application. The LASSO problem is defined by

$$\underset{\beta_0,\beta}{argmin} \sum_t \left(y^h_{t+h} - \beta_0 - \sum_i \beta_i x_{it} \right)^2 \text{ subject to } \sum_i |\beta_i| \leq s,$$

where $s \geq 0$ is a complexity parameter that determines the amount of shrinkage. The idea of penalizing the coefficients is that if there are so many correlated predictors, their coefficients are poorly estimated. This problem can be minimized by shrinking some coefficients to zero and imposing restrictions on the others.

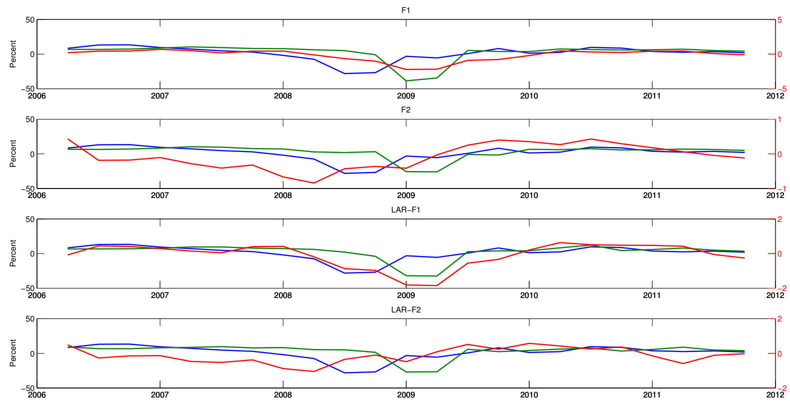

Figure 4: Diffusion indexes (red line, right scaled), average GDP growth over two-quarters (blue line, left scaled, at an annual rate) and the forecasts produced for the 2006-2011 period (green line, left scaled)

4.2 Graphical Results

In the remainder of this section, we closely follow the Stock and Watson (2003)'s analysis. We begin with a graphical analysis of forecasts for the 2006-2011 period. Graphs are useful for understanding which predictors start falling in advance of the 2009 fall in GDP. If it does so, it contains useful information for forecasting the 2009 decline in GDP.

Figures 4-6 show the results for the two-quarter-ahead forecasts, $h = 2$. The dates are the forecast dates. The red line shows the value of the predictors, the blue line shows the annualized value of GDP growth for two quarters ahead, and the green line shows the two-quarter-ahead forecast. The predictor's scale is given on the right axis while the scale for GDP growth and its forecasts is provided on the left axis.

The forecasts produced using the diffusion indexes are shown in Figures 4. The panels labeled as F1 and F2 are for the first and the second principal components, respectively. Those labeled as LAR-F1 and LAR-F2 correspond to the Bai and Ng (2008)'s modification, where we first select the 20 sectoral growth rates that are most correlated to the future GDP growth, after which we extract the principal components from this selected

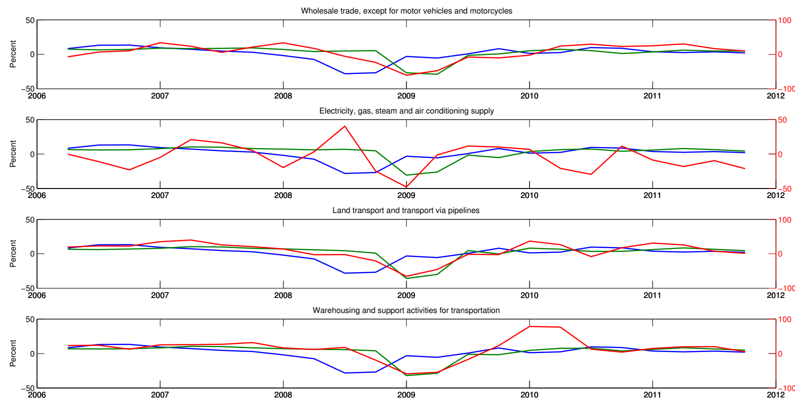

Figure 5: Average growth rate of selected hub sectors over two quarters (red line, right scaled, at an annual rate), average GDP growth over two-quarters, at an annual rate (blue line, left scaled, at an annual rate) and the forecasts produced for the 2006-2011 period (green line, left scaled)

set. We see that F2 and LAR-F2 fall in advance, which clearly signals the slowdown of the economy, while F1 and LAR-F1 do not show this slowdown.

The forecasts produced using the growth rates of selected sectors are presented in Figures 5 and 6. We see that the growth rate of the hub sectors and the sectors most correlated with the diffusion indexes only start falling coincident with the 2009 fall in GDP. In contrast, we see in Figure 6 that it is different for the growth of the sectors that make heavy use of imported inputs. In particular, the growth of *Chemical products* started falling in 2007Q1, providing a clear signal of future economic slowdown. Furthermore, a similar fall for *Rubber and plastic products* is occurred even earlier, in 2006Q3 but it increases from 2007Q3 to 2008Q1 and continues to fall further afterwards. These results suggest that the diffusion indexes and the growth of the sectors that make heavy use of imported inputs produce better forecasts, which is a result confirmed in the quantitative analysis provided in the next section.

In the appendix, we further present the forecasts for the horizons $h = 1,3$ and 4 in Figures 7-9. The graphical results for other horizons also suggest similar patterns in advance

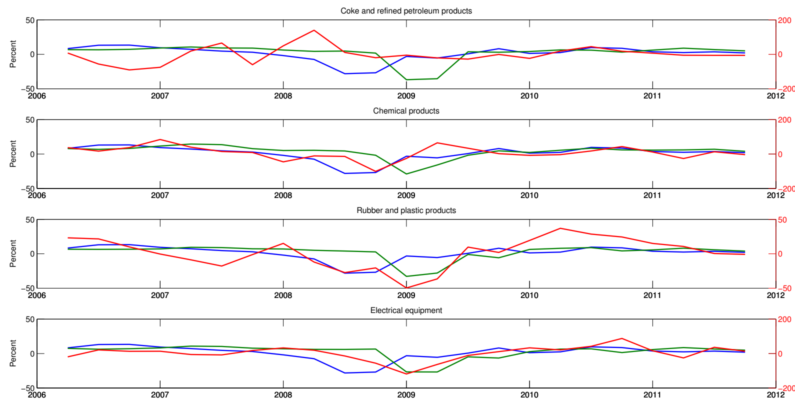

Figure 6: Average growth rate of the sectors heavily using imported inputs (red line, right scaled), average GDP growth over two-quarters, at an annual rate (blue line, left scaled, at an annual rate) and the forecasts produced for the 2006-2011 period (green line, left scaled)

of the 2009 fall in the case of LAR-F2 and the growth of some of the sectors that make heavy use of imported inputs.

4.3 Quantitative Analysis

We now turn to a quantitative analysis of the GDP growth forecasts for one to four quarters ahead (h = 1, 2, 3 and 4) for the 2006-2011 and 2012-2018 periods.

Our benchmark model is an autoregression. For comparison of the forecast i relative to the benchmark forecast, we use the relative mean squared error (MSE)

$$\text{relative MSE} = \frac{\sum_{t=T_1}^{T_2-h}\left(y_{t+h}^h - \hat{y}_{i,t+h}^h\right)^2}{\sum_{t=T_1}^{T_2-h}\left(y_{t+h}^h - \hat{y}_{0,t+h}^h\right)^2},$$

where $\hat{y}_{i,t+h}^h$ is the forecast from model i, $\hat{y}_{0,t+h}^h$ is the forecast from the benchmark autoregression, and T_1 and $T_2 - h$ are the first and the last forecast dates respectively. We set T_1 = 2006Q2 and T_2 = 2011Q4 + h once, and another time we set T_1 = 2012Q1 and T_2 = 2017Q3 + h. If the relative MSE is less than one, the forecast from model i performs better than the benchmark autoregression forecast.

The results are presented in Table 6. The first part of the table presents the root mean squared forecast error of the benchmark autoregression. The first period, including the 2009-fall in GDP, corresponds to much higher forecast errors with the annualized root mean squared forecast error amounting to 15.94 percent ($h = 1$), 13.97 percent ($h = 2$), 13.08 percent ($h = 3$) and 13.29 percent ($h = 4$). These numbers reduce in the second period to 3.55 percent, 2.70 percent, 2.42 percent and 2.62 percent, respectively. Inspection of the forecasts presented in Figures 4-6 shows that the forecasts generally do not perform well around the 2009-fall. The 2006-2011 root mean squared errors are very large relative to those presented for the 2012-2018 period. Excluding the four quarters 2009Q1-2009Q4 decreases them to 7.39 percent ($h = 1$), 5.36 percent ($h = 2$), 8.80 ($h = 3$) and 9.78 ($h = 4$).

The second part of the table shows the relative MSEs for the diffusion index forecasts (F1, F2, LAR-F1 and LAR-F2 denote the first principal component, the second principal component, the first and the second principal components extracted from the 20 sectoral rates selected by LAR, respectively), the forecasts produced by using the rates of selected sectors (D indicates that the sector's rate is most correlated with the factors, H indicates a hub sector, E indicates a sector with a large export, and I indicates a sector that makes heavy use of imported inputs), and the forecasts produced using a combination of individual sectoral rates and/or diffusion indexes (E1:4 denotes all the four sectors with large export and I1:4 denotes all four sectors using a large amount of imported inputs).

The results based on the diffusion index forecasts suggest that the best forecasts, outperforming the benchmark autoregression in the both subsample periods, are produced using LAR-F2 and LAR-F1:2. The MSE ratios from LAR-F2 are 0.89 ($h = 1$), 0.78 ($h = 2$), 0.72 ($h = 3$) and 0.81 ($h = 4$) in the first period. These numbers are 0.90, 0.84, 1.04 and 0.93 in the second period, respectively. The MSE ratios from LAR-F1:2 are 0.98 ($h = 1$), 0.7 ($h = 2$), 0.62 ($h = 3$) and 0.46 ($h = 4$) in the first period, and 0.74, 0.84, 0.89 and 1.01

in the second period, respectively. Overall, comparison of the accuracy of the F1, F2, and F1:2 forecasts with those made using LAR-F1, LAR-F2, and LAR-F1:2 implies that the relative accuracy of the forecasts improves in our sectoral data by carrying out the Bai and Ng (2008)'s modification. Furthermore, the best forecasts between the selected sectors are made using the growth of two of the top importing sectors (the *chemical products* sector and the *rubber and plastic products* sectors) which, similarly to LAR:F2 and LAR-F1:2, outperform the benchmark forecasts for most the horizons and in the both of the periods. For the case of the *chemical products* sector, the MSE ratios amount to 0.82 (h = 1), 0.66 (h = 2), 0.83 (h = 3) and 0.80 (h = 4) for the first period, which remain below one for three out of four horizons in the second period. The ratios for the case of the *rubber and plastic products* sector are 1.00 (h = 1), 0.89 (h = 2), 0.9 (h = 3) and 0.93 (h = 4) in the first period and 0.92 (h = 1), 0.85 (h = 2), 0.69 (h = 3) and 0.67 (h = 4) in the second period. The growth rates of other selected sectors do not produce better forecasts than the benchmark model. However, the *Land transport and transport via pipelines* sector, one of the hubs and one of the top exporting sectors, provides better forecasts for GDP growth for shorter horizons (h = 1 or 2).

Finally, since diffusion indexes and the growth of intermediate importers provide the most accurate forecasts, as a final exercise we combine them into a bigger set of predictors, which we in turn use to forecast GDP growth. The MSE ratios are presented in the final part of Table 6. The best forecasts are produced using the rates of all the four top importing sectors, I1:4, and also those produced using the rates of these four sectors together with LAR-F2 (labeled as LAR-F2 & I1:4). Considering the I1:4 forecasts, the MSE ratios are 0.88 (h = 1), 0.72 (h = 2), 0.81 (h = 3) and 0.83 (h = 4) in the first period and, in the second period, they are 0.89, 0.83, 0.66 and 0.55, respectively. These numbers for LAR-F2 & I1:4 are 0.88 (h = 1), 0.70 (h = 2), 0.64 (h = 3) and 0.59 (h = 4) in the first period, and 0.85, 0.87, 0.83, and 0.57 in the second period, respectively. Furthermore, note that the growth

Root mean squared error of the autoregression							
2006-2011				2012-2018			
h = 1	**h = 2**	**h = 3**	**h = 4**	**h = 1**	**h = 2**	**h = 3**	**h = 4**
15.94	13.97	13.08	13.29	3.55	2.70	2.42	2.62

	MSE ratios to the MSE of the autoregression							
	2006-2011				2012-2018			
	h = 1	**h = 2**	**h = 3**	**h = 4**	**h = 1**	**h = 2**	**h = 3**	**h = 4**
Diffusion Indexes								
F1	1.13	0.98	1.08	1.12	0.95	0.67	1.02	1.07
F2	1.03	**0.72**	**0.74**	**0.77**	0.90	0.81	1.06	1.02
F1:2	1.13	**0.83**	**0.85**	**0.86**	0.92	0.58	1.06	1.08
LAR-F1	1.07	**0.79**	**0.86**	**0.83**	0.78	0.97	0.88	1.11
LAR-F2	**0.89**	**0.78**	**0.81**	**0.72**	**0.90**	**0.84**	1.04	**0.93**
LAR-F1:2	0.98	**0.70**	**0.62**	**0.46**	**0.74**	**0.84**	**0.89**	1.01
Selected sectors								
Wholesale and retail trade and repair of motor vehicles (D)	0.82	0.83	0.96	1.06	1.08	0.70	0.95	0.99
Wholesale trade, except for motor vehicles (D, H, E)	0.99	**0.84**	**0.90**	**0.89**	0.94	1.26	1.16	1.12
Retail trade, except for motor vehicles (D)	1.02	1.01	1.07	1.12	0.93	1.03	1.23	1.04
Food and beverage service activities (D)	1.00	0.91	0.98	1.01	1.06	0.92	1.13	1.12
Electricity, gas, steam and air conditioning supply (H)	1.05	0.92	0.97	1.01	0.91	1.01	0.98	0.95
Land transport and transport via pipelines (H, E)	**0.95**	**0.93**	**0.94**	1.04	0.78	0.61	1.15	1.05
Warehousing and support activities for transportation (H)	0.96	0.91	1.01	0.97	0.96	1.01	1.02	1.03
Real estate activities (H)	1.02	1.03	0.98	1.02	0.98	1.02	1.09	1.12
Manufacturing of Beverages (E)	1.02	0.93	0.97	0.97	0.96	1.04	1.05	1.06
Coke and refined petroleum products (E, I)	1.00	0.99	0.92	1.02	0.85	0.89	1.03	1.01
Chemical products (I)	**0.82**	**0.66**	**0.83**	**0.80**	0.76	0.97	1.25	0.88
Rubber and plastic products (I)	1.00	**0.89**	**0.90**	**0.93**	**0.92**	**0.85**	**0.69**	**0.67**
Electrical equipment (I)	1.01	0.91	0.95	1.05	0.93	1.00	1.01	1.26
Combination of selected sectors and diffusion indexes								
E1:4	1.00	0.99	0.87	0.93	0.73	0.87	1.26	1.22
II:4	**0.88**	**0.72**	**0.81**	**0.83**	**0.89**	**0.83**	**0.66**	**0.55**
LAR-F1 & II:4	1.07	**0.71**	**0.94**	**0.75**	**0.71**	**0.77**	**0.61**	**0.72**
LAR-F2 & II:4	**0.88**	**0.70**	**0.64**	**0.59**	**0.85**	**0.87**	**0.83**	**0.57**

Table 6: The entries of the upper panel are root mean squared forecast errors for GDP growth, at an annual rate, produced by the benchmark autoregression. The remaining entries are the MSE ratios for each column horizon (h = 1, 2, 3 and 4). For each forecast, if MSE ratios are smaller that 0.95 for three out of four horizons in each subsample period, the MSE ratios are made bold. See text for description about forecast construction and the quantitative analysis carried out in this table.

in the sectors with larger exports contains much less interesting forecasting power for GDP growth compared to the growth in the sectors with larger imports. This is implied from the MSE ratios of the I1:4 forecasts which are, for most of the horizons, considerably smaller than the ratios of the E1:4 forecasts.

5 Summary and Conclusions

We describe a quarterly sectoral panel dataset combined with input-output tables for Lithuania, and use it to forecast GDP growth in the 2006-2011 and 2012-2018 periods. Our results show that, after 2009, there is a strong covariation among sectoral production growth rates. This degree of covariation can be explained by a few factors which are most correlated to the growth rates of the retail and wholesale sectors. We further produce real time forecasts for GDP growth, where we consider the diffusion index forecasts and the forecasts produced using the growth rates of selected individual sectors. Our results, compared to benchmark autoregression, suggest that the diffusion indexes and production growth of the sectors that make heavy use of imported inputs have a useful forecasting power for the growth rate of GDP in the 2006-2011 and 2012-2018 periods.

An interesting extension of our analysis would be to test the performance of the real time forecasts using vintage series for GDP and the sectors. Unfortunately, sectoral data was available to us only in revised form, but we think that this is a promising avenue for investigation.

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica 80*(5), 1977–2016.

Atalay, E. (2017). How important are sectoral shocks? *American Economic Journal: Macroeconomics 9*(4), 254–80.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics 146*(2), 304 – 317. Honoring the research contributions of Charles R. Nelson.

Boivin, J. and S. Ng (2005). Understanding and comparing factor-based forecasts. Technical report, National Bureau of Economic Research.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*(1), 169 – 194.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2018). The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy. *Review of Economic Studies 85*(4), 2042–2096.

Carvalho, V. M. (2014). From micro to macro via production networks. *Journal of Economic Perspectives 28*(4), 23–48.

Constantinescu, M. and K. Barauskaite (2018). Network-based macro fluctuations: what about an open economy? *Baltic Journal of Economics 18*(2), 95–117.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2004). Least angle regression. *The Annals of statistics 32*(2), 407–499.

Eickmeier, S. and C. Ziegler (2008). How successful are dynamic factor models at forecasting output and inflation? a meta-analytic approach. *Journal of Forecasting 27*(3), 237–265.

Forni, M. and L. Reichlin (1998). Let's get real: a factor analytical approach to disaggregated business cycle dynamics. *The Review of Economic Studies 65*(3), 453–473.

Gabaix, X. (2011). The granular origins of aggregate fluctuations. *Econometrica 79*(3), 733–772.

Long, J. B. J. and C. I. Plosser (1987). Sectoral vs. aggregate shocks in the business cycle. *The American Economic Review 77*(2), 333–336.

Lucas, R. E. (1977). Understanding business cycles. In *Carnegie-Rochester conference series on public policy*, Volume 5, pp. 7–29. Elsevier.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American statistical association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–162.

Stock, J. H. and M. W. Watson (2003). How did leading indicator forecasts perform during the 2001 recession? *Economic Quarterly* (Summer), 71–90.

Stock, J. H. and M. W. Watson (2016). Dynamic factor models, factor-augmented vector autoregressions, and structural vector autoregressions in macroeconomics. In *Handbook of macroeconomics*, Volume 2, pp. 415–525. Elsevier.

	Sector	NACE code
1	Wholesale and retail trade and repair of motor vehicles and motorcycles	G45
2	Wholesale trade, except of motor vehicles and motorcycles	G46
3	Retail trade, except of motor vehicles and motorcycles	G47
4	Food and beverage service activities	I56

Table 8: Turnover indexes of 4 retail and wholesale sectors

Appendix: List of Sectors

	Sector	NACE code
1	Extraction of crude petroleum and natural gas	B06
2	Other mining and quarrying	B08
3	Manufacture of beverages	C11
4	Manufacture of textiles	C13
5	Manufacture of wearing apparel	C14
6	Manufacture of leather and related products	C15
7	Manufacture of wood and of products of wood and cork, except for furniture; manufacture of articles of straw and plaiting materials	C16
8	Manufacture of paper and paper products	C17
9	Printing and reproduction of recorded media	C18
10	Manufacture of coke and refined petroleum products	C19
11	Manufacture of chemicals and chemical products	C20
12	Manufacture of basic pharmaceutical products and pharmaceutical preparations	C21
13	Manufacture of rubber and plastic products	C22
14	Manufacture of other non-metallic mineral products	C23
15	Manufacture of basic metals	C24
16	Manufacture of fabricated metal products, except machinery and equipment	C25
17	Manufacture of computer, electronic and optical products	C26
18	Manufacture of electrical equipment	C27
19	Manufacture of machinery and equipment n.e.c.	C28
20	Manufacture of motor vehicles, trailers and semi-trailers	C29
21	Manufacture of other transport equipment	C30
22	Manufacture of furniture	C31
23	Other manufacturing	C32
24	Repair and installation of machinery and equipment	C33
25	Electricity, gas, steam and air conditioning supply	D
26	Water supply; sewerage, waste management and remediation activities	E
27	Construction	F

Table 7: Production indexes of 27 manufacturing and construction sectors

	Sector	NACE code
1	Land transport and transport via pipelines	H49
2	Water transport	H50
3	Air transport	H51
4	Warehousing and support activities for transportation	H52
5	Postal and courier activities	H53
6	Accommodation	I55
7	Publishing activities	J58
8	Motion picture, video and television programme production, sound recording and music publishing activities	J59
9	Programming and broadcasting activities	J60
10	Telecommunications	J61
11	Computer programming, consultancy and related activities	J62
12	Information service activities	J63
13	Real estate activities	L68
14	Legal and accounting activities	M69
15	Activities of head offices; management consultancy activities	M70
16	Architectural and engineering activities; technical testing and analysis	M71
17	Advertising and market research	M73
18	Other professional, scientific and technical activities	M74
19	Veterinary activities	M75
20	Rental and leasing activities	N77
21	Employment activities	N78
22	Travel agency, tour operator reservation service and related activities	N79
23	Security and investigation activities	N80
24	Services to buildings and landscape activities	N81
25	Office administrative, office support and other business support activities	N82
26	Education	P
27	Human health and social work activities	Q
28	Arts, entertainment and recreation	R
29	Repair of computers, personal and household goods; other personal service activities	S95-S96

Table 9: Sale indexes of 29 services sectors

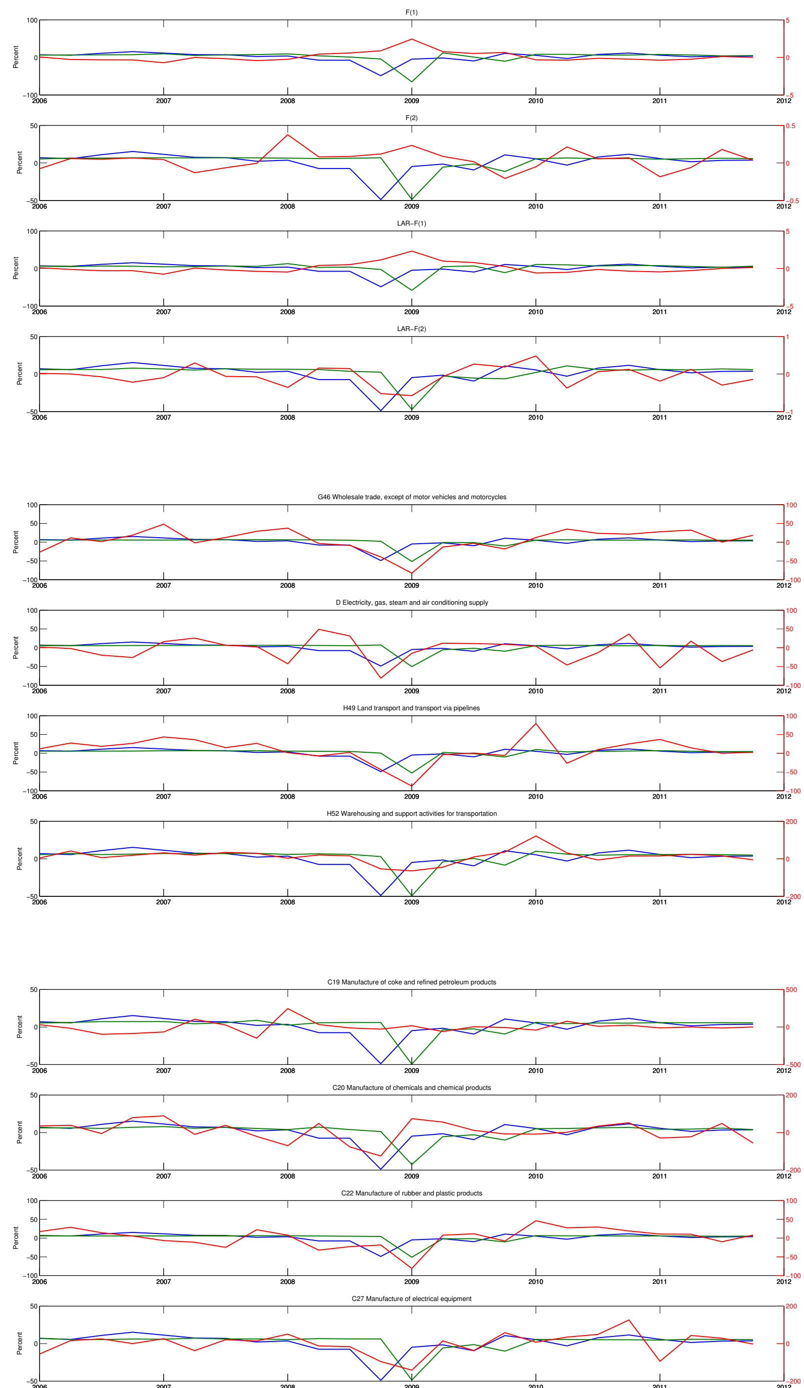

Figure 7: This figure shows the forecasts shown in Figures 4-6 for the horizon $h = 1$.

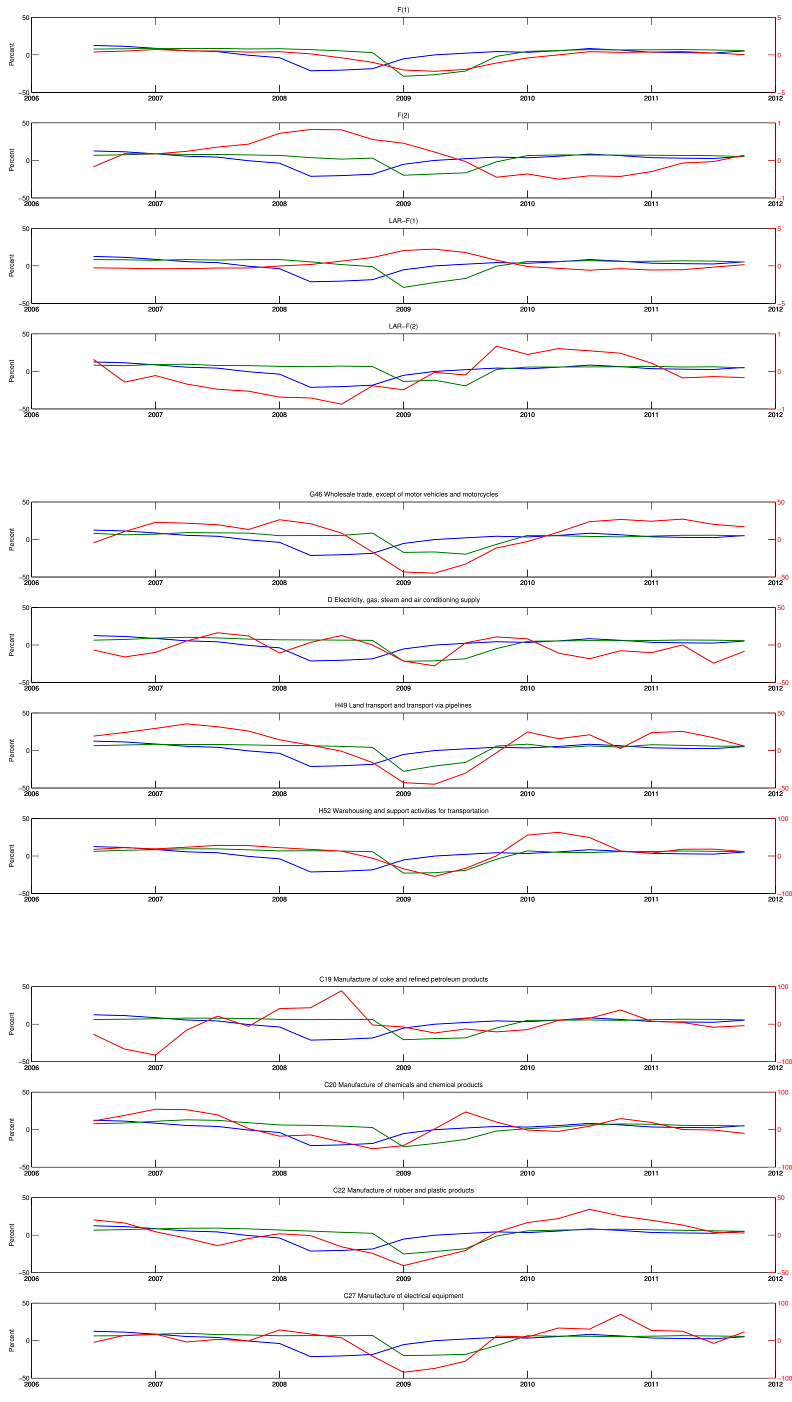

Figure 8: This figure shows the forecasts shown in Figures 4-6 for the horizon $h = 3$.

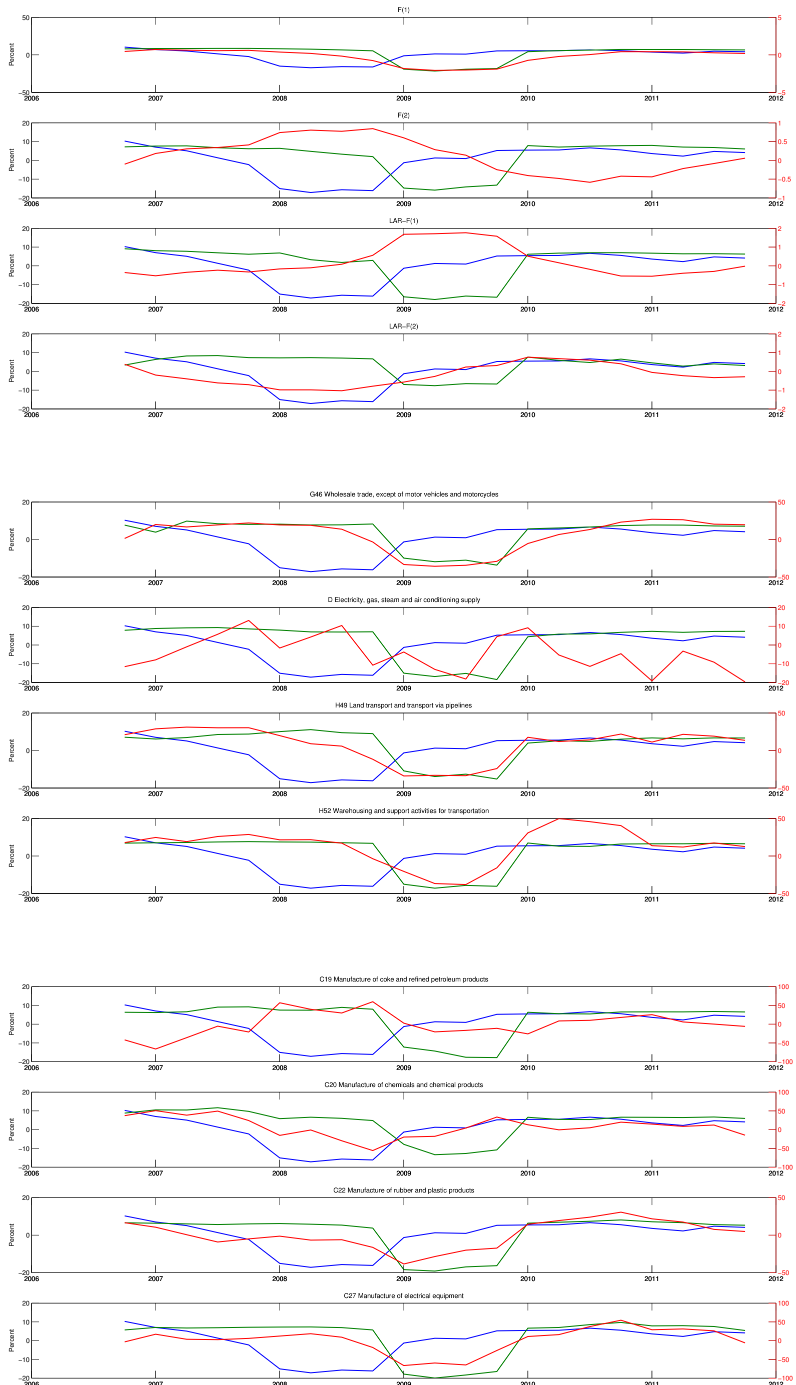

Figure 9: This figure shows the forecasts shown in Figures 4-6 for the horizon $h = 4$.